

# CARACTÉRISTIQUES GÉNÉRALES DES APPLICATIONS MODGEN – EXPLORATION DU MODÈLE RISKPATHS

Martin Spielauer

Statistique Canada – Division de la modélisation

Immeuble R.-H.- Coats, 24-O

Ottawa, K1A 0T6

[martin.spielauer@statcan.gc.ca](mailto:martin.spielauer@statcan.gc.ca)

RiskPaths est un modèle de microsimulation simple, à temps continu, basé sur les cas et à risques concurrentiels. Ce modèle est principalement utilisé comme outil d'enseignement, pour introduire la microsimulation à des spécialistes des sciences sociales et démontrer comment les modèles de microsimulation dynamique peuvent être programmés efficacement en utilisant le langage Modgen.

Modgen est un langage de programmation générique des modèles de microsimulation développé et maintenu à Statistique Canada.

RiskPaths, comme le langage de programmation Modgen et d'autres documents connexes, sont disponibles à [www.statcan.gc.ca/microsimulation/modgen/modgen-fra.htm](http://www.statcan.gc.ca/microsimulation/modgen/modgen-fra.htm)

## 1 Introduction

Modgen est un progiciel de développement de modèle de microsimulation développé et distribué par Statistique Canada. Il a été conçu pour permettre la création, la maintenance et la documentation de modèles de microsimulation sans devoir posséder des compétences avancées de programmation. Il permet d'utiliser de nombreuses approches de modélisation (modèles en temps continu ou discret, modèles orientés cas ou orientés temps, modèles généraux ou spécialisés, etc.). Modgen fournit aussi pour chaque modèle une interface visuelle commune qui met en œuvre des fonctions utiles, telles que la gestion des scénarios, l'entrée des paramètres, l'affichage des tableaux de sortie de l'exécution d'un modèle, la sortie graphique des biographies individuelles et l'affichage d'une documentation détaillée sur le modèle produite par Modgen.

Dans la discussion qui suit, nous présentons un simple modèle de microsimulation appelé RiskPaths qui a été mis en œuvre en utilisant Modgen. Nous commençons par une description des modèles statistiques sous-jacents, puis nous examinons des questions telles que celles de savoir ce que la microsimulation peut ajouter à l'analyse statistique initiale et quels sont les autres avantages qu'apporte la microsimulation à l'analyse globale. Ensuite, nous illustrons l'utilisation de certaines parties de l'interface visuelle de Modgen pour examiner les éléments du modèle RiskPaths.

RiskPaths, qui peut être utilisé comme modèle pour étudier l'absence d'enfants, a été développé à des fins de formation. Techniquement, il s'agit d'un modèle de cohorte à risques concurrents, démographiques, à un seul sexe (femmes uniquement), guidé par les données, spécialisé, en temps continu et orienté cas. Il s'appuie sur un ensemble de modèles de régression à risque constant par morceaux.

Essentiellement, RiskPaths permet de comparer le comportement démographique de base avant et après les transitions politiques et économiques vécues par la Russie et la Bulgarie autour de 1989. Ses paramètres ont été estimés d'après les données russes et bulgares de l'enquête sur les générations et le sexe menée autour de 2003-2004. La Russie et la Bulgarie constituent des études de cas intéressantes, puisqu'après l'effondrement du socialisme, ces deux pays ont connu les baisses de fécondité les plus importantes jamais observées dans l'histoire en temps de paix. De surcroît, les profils démographiques étaient très semblables et stables dans les deux pays durant la période de socialisme, ce qui permet de justifier l'utilisation de cohortes uniques comme moyen de comparaison (l'une représentant la vie à l'époque du socialisme et l'autre, celle d'une cohorte après la transition). De cette façon, le modèle nous permet de comparer le comportement démographique avant et après la transition, ainsi qu'entre les deux pays proprement dit.

## **2 RiskPaths : Les modèles statistiques sous-jacents**

### **2.1 Description générale**

Le modèle RiskPaths étant conçu pour étudier l'absence d'enfants, le principal événement est la première grossesse (qui est toujours supposé aboutir à une naissance). La grossesse peut survenir n'importe quand après le 15<sup>e</sup> anniversaire, le risque évoluant en fonction de l'âge ainsi que de la situation d'union. Les modèles statistiques sous-jacents sont des régressions à risque constant par morceaux. En ce qui concerne la fécondité, cela implique l'hypothèse d'un risque de grossesse constant pour un groupe d'âge (p. ex. 15 à 17,5 ans) et une situation d'union (p. ex. célibataire sans aucune union antérieure) donnés.

Pour les unions, nous distinguons quatre niveaux d'état possibles :

- célibataire;

- les trois premières années d'une première union;
- les années suivantes d'une première union;
- toutes les années d'une deuxième union.

(Après la dissolution d'une deuxième union, il est supposé que les femmes restent célibataires.)

Par conséquent, nous modélisons cinq événements d'union distincts :

- formation de la première union;
- dissolution de la première union;
- formation de la deuxième union;
- dissolution de la deuxième union;
- le changement de phase d'union qui survient après trois ans dans la première union.

Le dernier événement (changement de phase de l'union) est un événement horloge, qui diffère des autres événements en ce sens que le moment où il survient n'est pas stochastique mais prédéfini. (Un autre événement horloge dans le modèle est le changement de l'indice d'âge, tous les 2,5 ans.) Outre les unions et la fécondité, nous modélisons la mortalité — une femme peut mourir à n'importe quel point dans le temps. Nous arrêtons la simulation de la grossesse et des événements d'union quand une femme est décédée ou qu'elle est enceinte (car nous cherchons seulement à étudier l'absence d'enfants) ou à son 40<sup>e</sup> anniversaire, parce que les premières grossesses après cet âge sont très rares en Russie et en Bulgarie et ne sont donc pas prises en compte dans le présent modèle.

À 15 ans, une femme devient exposée à la fois au risque de grossesse et de formation d'une union. Il s'agit de risques concurrents. Nous tirons des durées aléatoires jusqu'à la première grossesse et jusqu'à la formation de la première union. À ce stade il existe deux autres événements concurrents — la mortalité et le changement de groupe d'âge. (Comme nous supposons que les risques de grossesse et de formation d'une union évoluent avec l'âge, les risques qui sous-tendent les durées aléatoires ne s'appliquent qu'à une période donnée — 2,5 ans dans notre modèle — et doivent être recalculés à chaque point dans le temps.)

Autrement dit, le 15<sup>e</sup> anniversaire sera suivi par l'un des quatre événements possibles suivants :

- la femme décède;
- elle devient enceinte;
- elle entre dans une union;
- elle entre dans un nouveau groupe d'âge à 17,5 ans, parce qu'aucun des trois événements qui précèdent n'a eu lieu avant qu'elle n'atteigne 17,5 ans.

Le décès ou la grossesse mettent fin à la simulation. Un indice de changement d'âge requiert que les temps d'attente pour les événements concurrents de formation d'une union et de grossesse soient mis à jour. L'événement de formation d'une union modifie le risque d'une première grossesse (le rendant beaucoup plus élevé) et change l'ensemble de risques concurrents. La femme n'est désormais plus exposée au risque de formation d'une première union, mais devient exposée au risque de dissolution de l'union.

## 2.2 Événements et estimations des paramètres

### 2.2.1 Première grossesse

Comme nous l'avons décrit plus haut, la première grossesse est modélisée au moyen d'un risque à un âge de référence et des risques relatifs dépendant de la situation d'union et de la durée. Le tableau 1 qui suit donne les estimations des paramètres pour la Bulgarie et la Russie avant et après la transition politique et économique.

**Tableau 1 : Risques d'une première grossesse**

	Bulgarie	Russie
15 à 17,5 ans	0.2869	0.2120
17,5 à 20 ans	0.7591	0.7606
20 à 22,5 ans	0.8458	0.8295
22,5 à 25 ans	0.8167	0.6505
25 à 27,5 ans	0.6727	0.5423
27,5 à 30 ans	0.5105	0.5787
30 à 32,5 ans	0.4882	0.4884
32,5 à 35 ans	0.2562	0.3237
35 à 37,5 ans	0.2597	0.3089
37,5 à 40 ans	0.1542	0.0909

	Avant la transition de 1989		Dix ans après la transition : 1999+	
	Bulgarie	Russie	Bulgarie	Russie
Pas dans une union	0.0648	0.0893	0.0316	0.0664
Trois premières années de la première union	1.0000	1.0000	0.4890	0.5067
Première union après trois ans	0.2523	0.2767	0.2652	0.2746
Deuxième union	0.8048	0.5250	0.2285	0.2698

Les données du tableau 1 sont interprétées de la façon suivante dans le modèle. Aussi longtemps qu'une femme n'a pas formé une union, nous devons multiplier son risque de base variant en fonction de l'âge d'une première grossesse par le risque relatif de « ne pas être dans une union ».

Par exemple, nous pouvons calculer le risque de grossesse d'une femme célibataire de 20 ans de la cohorte bulgare avant la transition comme étant  $0,8458 * 0,0648 = 0,05481$ . À ce taux de  $\lambda=0,05481$  :

- le temps d'attente moyen prévu de l'événement de grossesse est  $1/\lambda = 1/0,05481 = 18,25$  ans;
- la probabilité qu'une femme ne devienne pas enceinte dans les 2,5 ans qui suivent (étant donné qu'elle demeure célibataire) est  $\exp(-\lambda t) = \exp(-0,05481 * 2,5) = 87,2$  %.

Donc, à son 20<sup>e</sup> anniversaire, nous pouvons tirer une durée aléatoire jusqu'à la première grossesse sous forme d'un nombre aléatoire suivant une loi uniforme (un nombre qui peut prendre n'importe quelle valeur entre 0 et 1 avec la même probabilité) en utilisant la formule :

$$\text{Durée aléatoire} = -\ln(\text{RandomUniform}) / \lambda.$$

Comme nous l'avons calculé plus haut, dans 87,2 % des cas, aucune conception n'aura lieu au cours des 2,5 années qui suivent. Par conséquent, si nous tirons un nombre aléatoire suivant une loi uniforme plus petit que 0,872, le temps d'attente correspondant sera plus long que 2,5 ans, puisque  $-\ln(\text{RandomUniform}) / \lambda = -\ln(0,872)/0,05481 = 2,5$  ans. Un tirage aléatoire supérieur à 0,872 résultera en un temps d'attente plus faible que 2,5 ans — dans cette situation, si la femme n'entre pas dans une union avant l'événement de grossesse, la grossesse a lieu dans notre simulation.

Afin de poursuivre cet exemple, supposons que le premier événement qui se produit dans notre simulation est la formation d'une union à l'âge de 20,5 ans. Nous devons maintenant mettre à jour le risque de grossesse. Tandis que le risque de base reste le même pour les deux années qui suivent (c.-à-d. 0,8458), le risque relatif est maintenant égal à 1,0000 (conformément à la catégorie de référence du tableau 1) parce que la femme se trouve dans les trois premières années d'une union. Le nouveau taux de risque de grossesse (applicable aux deux années suivantes, jusqu'à l'âge de 22,5 ans) est maintenant considérablement plus élevé, soit  $0,8458 * 1,0000 = 0,8458$ . Le temps d'attente moyen à ce taux n'est donc que de  $1/0,8458 = 1,18$  an et, pour tout nombre aléatoire supérieur à  $\exp(-0,8458 * 2) = 0,1842$ , le temps d'attente simulé sera inférieur à deux ans. Autrement dit, 81,6 % ( $1 - 0,1842$ ) de femmes connaîtront une première grossesse dans les deux premières années d'une première union ou d'un premier partenariat qui a débuté à l'âge de 20,5 ans.

### **2.2.2 Formation de la première union**

Les risques sont donnés sous forme de taux constants par morceaux évoluant avec l'âge. De nouveau, nous modélisons des intervalles d'âge de 2,5 ans. Voici les taux pour les femmes avant toute conception, puisqu'un tel événement mettrait fin à notre simulation.

**Tableau 2 : Risques de formation de la première union**

	Avant la transition de 1989		Dix ans après la transition : 1999+	
	Bulgarie	Russie	Bulgarie	Russie
15 à 17,5 ans	0.0309	0.0297	0.0173	0.0303
17,5 à 20 ans	0.1341	0.1342	0.0751	0.1369
20 à 22,5 ans	0.1672	0.1889	0.0936	0.1926
22,5 à 25 ans	0.1656	0.1724	0.0927	0.1758
25 à 27,5 ans	0.1474	0.1208	0.0825	0.1232
27,5 à 30 ans	0.1085	0.1086	0.0607	0.1108
30 à 32,5 ans	0.0804	0.0838	0.0450	0.0855
32,5 à 35 ans	0.0339	0.0862	0.0190	0.0879
35 à 37,5 ans	0.0455	0.0388	0.0255	0.0396
37,5 à 40 ans	0.0400	0.0324	0.0224	0.0330

L'exemple de paramétrisation donné au tableau 2 s'interprète de la façon suivante : le risque de formation d'une première union des femmes bulgares de la première cohorte est nul jusqu'au 15<sup>e</sup> anniversaire; après cela, il évolue par intervalle de 2,5 ans de 0,0309 à 0,1341, puis de 0,1341 à 0,1672, et ainsi de suite. Le risque le plus élevé s'observe pour le groupe des 20 à 22,5 ans — au taux de 0,1672, la durée prévue jusqu'à la formation d'une union est de  $1/0,1672=6$  ans. Une femme qui est célibataire à son 20<sup>e</sup> anniversaire a une probabilité de 34 % de former une première union au cours des 2,5 années suivantes ( $p=1-\exp(-0,1672*2,5)$ ).

### 2.2.3 Formation de la deuxième union

Une femme devient exposée au risque de formation d'une deuxième union si et quand sa première réunion est dissoute. Contrairement à la formation de la première union qui est fondée sur l'âge, ce processus ne démarque pas à un point fixe dans le temps, mais est déclenché par un autre événement (dissolution de la première union). Par conséquent, les intervalles de temps des taux de risque constants par morceaux estimés ont trait au temps écoulé depuis la dissolution de la première union.

**Tableau 3 : Risques de formation de la deuxième union**

	Avant la transition de 1989		Dix ans après la transition : 1999+	
	Bulgarie	Russie	Bulgarie	Russie
<2 ans après la dissolution	0.1996	0.2554	0.1457	0.2247
2 à 6 ans après la dissolution	0.1353	0.1695	0.0988	0.1492
6 à 10 ans après la dissolution	0.1099	0.1354	0.0802	0.1191
10 à 15 ans après la dissolution	0.0261	0.1126	0.0191	0.0991
>15 ans après la dissolution	0.0457	0.0217	0.0334	0.0191

#### 2.2.4 Dissolution d'une union

La première ainsi que la deuxième union peut être dissoute, le processus débutant au moment de la formation de la première et de la deuxième union, respectivement. Comme la taille de l'échantillon est très faible pour la modélisation de l'événement de dissolution de la deuxième union, nous ne faisons pas la distinction entre les cohortes avant et après la transition pour cet événement.

Tableau 4 : Risques de dissolution de la première union

	Avant la transition de 1989		Dix ans après la transition : 1999+	
	Bulgarie	Russie	Bulgarie	Russie
Première année de l'union	0.0096	0.0380	0.0121	0.0601
Durée de l'union de 1 à 5 ans	0.0200	0.0601	0.0252	0.0949
Durée de l'union de 5 à 9 ans	0.0213	0.0476	0.0269	0.0752
Durée de l'union de 9 à 13 ans	0.0151	0.0408	0.0190	0.0645
Durée de l'union >13 ans	0.0111	0.0282	0.0140	0.0445

Tableau 5 : Risques de la dissolution de la deuxième union

	Bulgarie	Russie
Trois premières années de l'union	0.0371	0.0810
Durée de l'union de 3 à 9 ans	0.0128	0.0744
Durée de l'union de 9 ans et plus	0.0661	0.0632

#### 2.2.5 Mortalité

Dans ce modèle échantillon, nous laissons à l'utilisateur le soin de préciser les probabilités de décès selon l'âge ou de « désactiver » la mortalité, ce qui permet d'étudier la fécondité sans l'interférence de la mortalité. Dans ce deuxième cas, toutes les femmes atteignent l'âge maximal de 100 ans. Si l'utilisateur choisit de simuler la mortalité, les probabilités spécifiées sont converties par le modèle en taux de risque constants par morceaux (en se basant sur la formule  $-\ln(1-p)$  pour  $p < 1$ ) de sorte que le décès peut survenir à n'importe quel moment durant une année. Si la probabilité est fixée à 1 (comme cela est le cas quand âge = 100), il est supposé que le décès survient immédiatement.

### 3 Qu'attendons-nous du modèle de microsimulation RiskPaths?

#### 3.1 Que peut ajouter la simulation à l'analyse statistique?

Avant de pouvoir répondre à la question de savoir ce que la simulation ajoute à l'analyse statistique, nous devons bien comprendre ce que les résultats statistiques présentés à la section

précédente révèlent. Les résultats des estimations pour les deux pays et pour les deux cohortes nous permettent d'étudier les similarités et les différences entre ces pays, ainsi que l'évolution des paramètres au cours du temps *séparément* pour chaque processus individuel. Nous constatons une similarité remarquable des paramètres entre les deux pays, surtout pour les cohortes d'avant la transition. Fondamentalement, la Bulgarie ne diffère de la Russie qu'en ce qui concerne le risque trois fois plus faible de dissolution d'une union et la vitesse plus lente de formation d'une deuxième union. Par conséquent, si nous comparons les cohortes avant et après la transition, nous observons des changements spectaculaires dans la plupart des processus. Le risque d'une première naissance a été réduit de moitié durant les trois premières années de la première union, sans redressement ultérieur, quoique les paramètres aient changé relativement peu après trois années dans une union. En outre, dans les deuxièmes unions, la fécondité a baissé de plus de 50 %. La différence la plus importante entre les deux pays après la transition s'observe pour la formation d'une première union, les taux ayant été réduits de moitié en Bulgarie, mais étant demeurés stables en Russie. Pour la dissolution de la première union nous observons le tableau opposé — le risque de dissolution d'une union a augmenté d'environ 40 % en Russie, mais est resté presque le même en Bulgarie.

Il s'agit là d'exemples typiques des éclaircissements que peut fournir l'*analyse d'un processus unique*. Nous avons décomposé un système complexe en ses processus constitutifs et étudié les changements à l'intérieur de ces processus. Dans le cas de la fécondité, nous avons introduit des risques relatifs — nous avons étudié comment certains facteurs (ici différentes situations d'union) influencent un processus *unique*. Il s'agit d'une question analytique très typique et ce genre de recherche abonde dans la littérature scientifique.

La puissance de la microsimulation se dégage quand nous étudions divers processus simultanément. Même dans notre exemple démographique très simple, les résultats sont difficiles à interpréter quand nous nous intéressons à l'effet qu'ont sur les résultats agrégés les changements dans des processus uniques. Par exemple, quel est l'effet de l'accroissement de 40 % du risque de dissolution d'une union sur le fait de ne pas avoir d'enfants en Russie? Il dépendra de la fécondité en dehors des unions et dans les deuxièmes unions, ainsi que de la vitesse de formation d'une deuxième union. Le risque relatif de fécondité est plus élevé dans les deuxièmes unions qu'après trois années dans la première union, mais la formation de la deuxième union prend du temps (durant lequel la fécondité est très faible) et toutes les femmes ne forment pas une deuxième union. Ces effets s'annulent-ils les uns les autres ou la dissolution d'une union affecte-t-elle la fécondité, et dans quelle direction? Ce genre de question nous invite à utiliser la microsimulation pour une *analyse de sensibilité*. Comment agrégeons-nous le changement de résultat en réponse à la variation d'un seul paramètre? Notons que nous sommes maintenant passés d'une analyse au niveau d'un processus unique à une *analyse du comportement d'un système*.



Une comparaison des deux cohortes suscite un type supplémentaire d'analyse de système — quelle est la contribution relative du changement dans les processus uniques aux résultats agrégés? Si nous comparons les deux cohortes simulées, nous voyons que l'absence d'enfants a augmenté considérablement dans les deux pays, mais encore plus en Bulgarie qu'en Russie. Nous pouvons utiliser la microsimulation pour *décomposer* les contributions des changements dans les divers processus au changement agrégé. Comment aurait évolué la prévalence de l'absence d'enfants si seuls les paramètres de fécondité avaient changé? Quelle est la contribution des changements concernant la formation d'une union? L'accroissement du risque de dissolution d'une union a-t-il contribué à l'accroissement de l'absence d'enfants en Russie? Naturellement, le changement agrégé n'est pas égal à la simple somme arithmétique des effets partiels. Certains changements de processus pourraient avoir un effet plus prononcé ou plus faible en présence de changements dans les autres processus. Par exemple, l'effet du changement de fécondité dans les deuxièmes unions dépendra fortement de la probabilité d'être dans une deuxième union, laquelle est conditionnelle aux risques de formation et de dissolution d'une première union. La microsimulation peut nous aider à déceler et à mieux comprendre ce genre d'interaction.

Quand nous examinons la cohorte après la transition, nous sommes déjà entrés dans le domaine des *prédictions*. Comme les données ont été recueillies 14 ans après la transition, en réalité, aucune cohorte postérieure à la transition n'a vécu l'entièreté de sa période reproductive. Donc, pour des mesures de cohorte comme l'absence d'enfants, l'évaluation de la cohérence avec d'autres sources de données est limitée à une comparaison à d'autres projections. Toutefois, nous pouvons aussi utiliser notre modèle pour refaire des prédictions sous diverses hypothèses quant aux futurs changements dans les processus. Nous pourrions avoir conçu une théorie qui nous mène à l'hypothèse que seules des parties des changements observés sont de nature permanente (p. ex. causés par un changement culturel), tandis que d'autres sont transitoires (p. ex. résultant d'une crise économique, par conséquent réversibles avec la reprise économique). Que se passerait-il si les taux de fécondité revenaient à leur valeur initiale alors que la formation plus lente d'une union (ultérieure) persiste, ou inversement? Ce genre d'analyse peut produire des résultats surprenants, car ce n'est pas nécessairement le renversement du processus qui, initialement, a eu l'effet global le plus important qui produira l'effet opposé le plus grand.

Existe-t-il des incidences en ce qui concerne les politiques? Bien que notre modèle soit évidemment trop simple pour l'analyse des politiques, il ne faut pas beaucoup d'imagination pour voir comment la microsimulation peut appuyer l'élaboration des politiques.

- Dans de nombreux cas, les phénomènes étudiés ont un rapport direct avec les politiques. La baisse de la fécondité aura par exemple une incidence sur la viabilité des systèmes de sécurité sociale. Un bon modèle de projection démographique peut par conséquent produire des données d'entrée précieuses pour une planification subséquente de bonne qualité. La

microsimulation est l'outil qui permet de combiner des modèles statistiques distincts en modèles de projection.

- Les événements simulés dans un modèle de microsimulation peuvent également être eux-mêmes les cibles des politiques. Un gouvernement pourrait chercher à influencer sur la fécondité, ce qui est possible s'il met en place des politiques capables d'influencer les processus modélisés. Cependant, nous devons d'abord arriver à comprendre la contribution individuelle de ces processus aux résultats agrégés que nous voulons modifier; donc, nous avons besoin de microsimulations. Si nous pouvons mettre un prix sur ces politiques, nous serons également capables d'utiliser la microsimulation pour trouver la combinaison de politiques la plus rentable, et donc d'étudier les effets secondaires possibles. (Sous le régime socialiste, la Russie et la Bulgarie avaient effectivement en place un ensemble de politiques puissantes, telles que l'impôt sur le célibat et l'accès privilégié au logement pour les jeunes couples mariés. Le prix de la régulation des choix de vie individuels s'est avéré être assez élevé.)
- La microsimulation nous permet de compléter les modèles résultant de l'analyse statistique par des scénarios stratégiques et des modèles de comptabilité économique détaillés. Elle offre un outil très naturel pour la simulation des politiques, car celles-ci sont définies au niveau individuel ou micro, ce qui aboutit à des applications qui intègrent la modélisation démographique et économique.

## **3.2 Caractéristiques souhaitées d'un modèle de microsimulation RiskPaths**

### **3.2.1 Entrées : Tableaux des paramètres, scénarios et paramètres de la simulation**

Même s'il est très simple, le modèle RiskPaths possède environ 130 valeurs de paramètre que les utilisateurs doivent être capables d'attribuer et de mémoriser comme il convient. Ces paramètres devraient être présentés de manière bien structurée dans l'application de microsimulation, sous forme de tableaux étiquetés, d'accès (ou de navigation) facile, qui peuvent être lus ou modifiés au besoin.

Quand nous utilisons un modèle, nous créons habituellement différents scénarios, c'est-à-dire différentes paramétrisations du modèle. Nous devons pouvoir sauvegarder ces scénarios afin que certaines simulations puissent être reproduites dans l'avenir. Les scénarios contiennent tous les tableaux des paramètres et, idéalement, des descriptions textuelles ou des notes supplémentaires qui décrivent les changements spécifiques intégrés dans chaque scénario. En outre, les scénarios doivent inclure les paramètres du scénario, tels que le nombre de cas simulés (étant donné que RiskPaths est un modèle orienté cas). Une grande taille d'échantillon réduira la variation Monte Carlo, mais au prix de l'exécution plus lente des simulations. Si nous nous intéressons uniquement aux agrégats généraux, des échantillons de plus petite taille pourraient suffire. Par

ailleurs, l'analyse détaillée d'événements rares ou une ventilation détaillée des résultats (p. ex. selon le groupe d'âge) nécessitera de grands échantillons. En outre, l'utilisateur pourrait ne pas vouloir produire toutes les sorties disponibles. Limiter la sortie souhaitée peut de nouveau accélérer les simulations, mais aussi donner lieu à une présentation plus concise et focalisée des résultats, conformément aux besoins de l'utilisateur.

Tous les éléments susmentionnés (tableaux des paramètres, notes descriptives, nombre de cas, choix de la sortie à produire) font partie d'un scénario. Dans nos applications RiskPaths, toute cette information devrait être mémorisée ensemble pour un scénario particulier et elle devrait pouvoir être extraite, consultée et modifiée facilement.

### **3.2.2 Sortie et visualisation des sorties**

Les modèles de microsimulation peuvent produire des sorties à deux niveaux, micro et macro. Une application de microsimulation pourrait en théorie écrire toutes les caractéristiques de niveau individuel et tous leurs changements au cours du temps dans un fichier et laisser à l'utilisateur la tâche d'analyser le fichier de données résultant au moyen d'un logiciel statistique. Dans le cas de RiskPaths, cela aboutirait à un fichier contenant les dates de tous les événements simulés qui surviennent au cours de la trajectoire de vie simulée de chaque individu. Six événements seulement peuvent avoir lieu dans une vie simulée, de sorte que chaque enregistrement de données contiendrait au plus six variables : quatre événements de formation/dissolution d'union, la grossesse et le décès. Toutefois, pour des applications plus complexes, la taille et la complexité du fichier pourraient être énormes.

En plus d'un fichier longitudinal de ce genre, nous pourrions aussi souhaiter une sortie transversale, dans laquelle sont enregistrés les états de tous les individus à un point particulier dans le temps. L'utilisation d'un tel fichier serait relativement limitée dans la simulation d'une seule cohorte, mais il ressemblerait à une enquête transversale ou à un recensement de population dans un modèle de population.

Habituellement, un utilisateur de modèle ne s'intéresse pas aux fichiers de microdonnées proprement dits, mais à leur analyse. Il agrégera généralement les données et produira des indicateurs et des tableaux sommaires. Si les développeurs d'un modèle savent comment les données simulées seront ou devraient être analysées, ces mesures et tableaux peuvent déjà être calculés et produits dans une exécution de l'application de microsimulation. Dans ce cas, les utilisateurs n'auront pas besoin d'exécuter des routines statistiques supplémentaires; ils pourront voir les résultats immédiatement après l'exécution d'une simulation. Dans notre modèle RiskPaths, la sortie ne comporte qu'un petit nombre de tableaux et d'indicateurs sommaires que nous nous attendons à être produits dans l'application. Nous nous intéressons au taux de

fécondité par âge, à l'absence d'enfants, à l'âge moyen au moment de la première grossesse, à la première grossesse selon la situation d'union et à certaines mesures de mortalité.

Comme pour les tableaux de paramètres, la sortie agrégée du modèle requiert une structure. Nous pourrions souhaiter présenter certaines mesures sommaires d'un ou plusieurs comportements apparentés regroupés dans un tableau et nous voudrions certainement ordonner le contenu du tableau de manière significative. En outre, comme pour les paramètres, les résultats tabulaires devraient être étiquetés de manière qu'ils soient faciles à lire et à comprendre.

Comme tous les résultats de microsimulation sont sujets à la variation Monte Carlo, les chiffres agrégés ne sont qu'une visualisation des résultats. Nous pourrions aussi vouloir obtenir des renseignements sur la distribution de chaque valeur du tableau. Ce genre d'information nous aiderait à sélectionner une taille de population suffisante pour un niveau de précision souhaité des résultats.

L'affichage graphique des trajectoires individuelles est un type particulier de sortie de microdonnées. Il peut s'agir d'une fonction utile, car elle donne aux utilisateurs une vue des individus simulés, donc un moyen de voir le fonctionnement des modèles statistiques. Elle peut aussi être utile aux développeurs des modèles car elle facilite le débogage. Puisque RiskPaths est un outil de formation, nous souhaitons montrer comment les biographies individuelles résultent des processus statistiques. Donc, outre les événements qui surviennent au cours de la vie, nous pourrions aussi vouloir voir comment les risques des autres événements évoluent au cours du temps et des situations au cours de la vie.

### **3.2.3 Interface utilisateur et documentation**

Jusqu'à présent, nous avons formulé des attentes concernant le contenu, l'affichage et la présentation des données d'entrée et de sortie du modèle. Du point de vue de l'utilisateur, devons-nous simplement ajouter un bouton de démarrage pour compléter l'application de microsimulation? Est-ce que toutes les applications logicielles contemporaines contiennent des fichiers d'aide. En tant qu'utilisateur de modèles de microsimulation, nous devrions nous attendre à avoir accès à une aide détaillée en ligne, non seulement concernant l'utilisation du logiciel de modélisation proprement dit, mais aussi au sujet des éléments particuliers du modèle et des interdépendances entre ces éléments.

## **4 Exploration de l'application RiskPaths de Modgen**

La suite de notre discussion comprend une description rapide de l'interface visuelle fournie par Modgen pour le modèle RiskPaths. Pour exécuter RiskPaths, il faut que l'application des composants préalables à Modgen (Modgen Prerequisites) ainsi que l'exécutable RiskPaths soient installés sur votre ordinateur. Comme toutes les applications Modgen, RiskPaths contient un

système d'aide qui contient de la documentation sur l'interface utilisateur (indépendante du modèle) de Modgen et sur le modèle RiskPaths proprement dit. Par conséquent, dans la description qui suit, nous nous concentrons sur les étapes centrales de l'exécution de RiskPaths, et nous vous laissons le soin d'explorer en détail le modèle et le logiciel en vous aidant des fichiers d'aide détaillés.

#### **4.1 L'interface utilisateur**

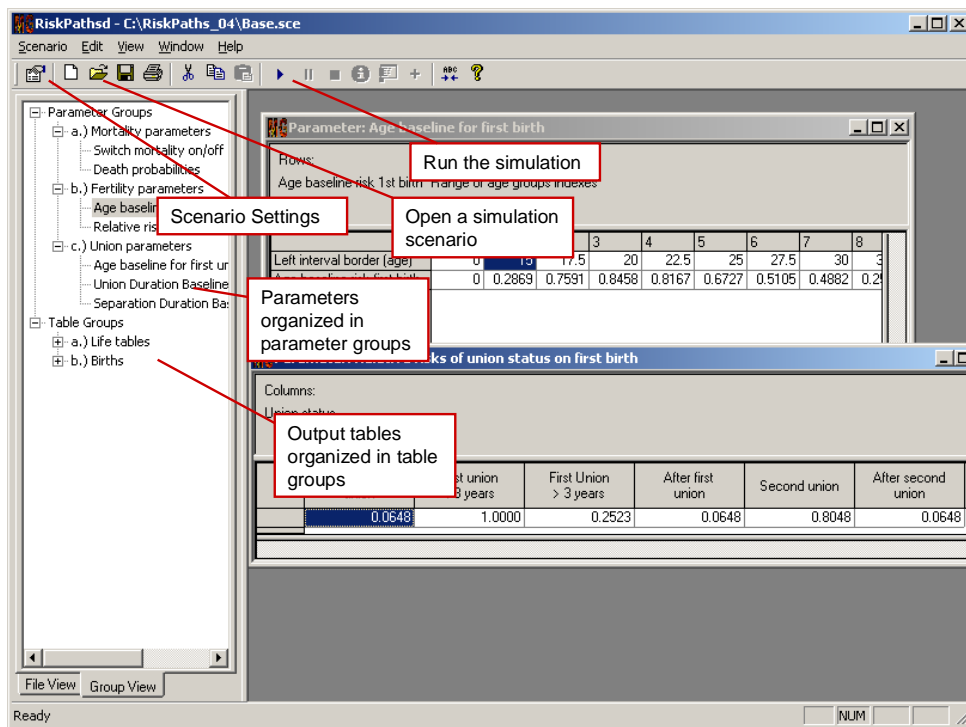
Toutes les applications Modgen ont la même interface utilisateur graphique (figure 1) qui comprend les éléments suivants :

- une barre de menu et une barre d'outils pour administrer et exécuter des scénarios, ainsi que pour obtenir de l'aide;
- une fenêtre de sélection contenant une liste groupée hiérarchiquement de tous les paramètres et tableaux de sortie du modèle;
- un cadre dans lequel tous les paramètres ou tableaux correspondants peuvent être affichés.

Au moment du démarrage de l'application RiskPaths.exe, la fenêtre de sélection et le cadre des tableaux sont vides, parce que nous devons d'abord charger (ou créer) un scénario de simulation. Pour cela, suivez les étapes suivantes :

- Ouvrez le scénario de simulation « base.sce ». Vous pouvez le faire en cliquant sur le bouton « Open » ou en choisissant « Open... » dans le menu « Scénario ».
- Choisissez les paramètres du scénario. Vous pouvez avoir accès à la boîte de dialogue concernant les paramètres en cliquant sur le bouton « Settings » ou en choisissant « Settings... » dans le menu « Scénario ». Spécifier un petit nombre de cas à simuler (p. ex. 10 000), afin que votre premier modèle soit exécuté rapidement. En outre, veiller à ce que « MS Access tracking » soit activé. Cela vous permettra de voir les bibliographies individuelles en utilisant l'outil BioBrowser qui vient avec Modgen.
- Sauvegardez votre scénario sous un nouveau nom en sélectionnant « Save as... » dans le menu « Scénario ».

**Figure 1 : Application RiskPaths**



## 4.2 Tableaux des paramètres

Les utilisateurs d'une application Modgen peuvent contrôler tous les paramètres contenus dans les tableaux de paramètres du modèle. Un tableau de paramètres individuel peut être choisi en cliquant sur son entrée dans la liste dans la fenêtre de sélection. Le tableau s'affiche alors dans le cadre d'affichage dans lequel il peut alors être édité. Les tableaux de paramètres Modgen peuvent avoir n'importe quel nombre de dimensions, variant d'un paramètre avec une case à cocher unique à des paramètres possédant de nombreuses caractéristiques ou dimensions (p. ex. région, sexe, âge, période).

**Figure 2 : Paramétrisation des risques de formation d'une première union**

	0	1	2	3	4	5	6
Left interval border (age)	0	15	17.5	20	22.5	25	27.5
Age baseline risk first union formation	0	0.030898	0.134066	0.167197	0.165551	0.147390	0.108470

## 4.3 Exécution d'une simulation

Cliquez sur le bouton « Run/resume » ou choisissez « Run/resume » dans le menu « Scénario ». L'état d'avancement de la simulation est affiché dans une boîte de dialogue indiquant les progrès. Dans le cas d'un petit échantillon de 10 000 acteurs, l'exécution prend environ 20 secondes.

Quand l'exécution du modèle est achevée, tous les tableaux de sortie ont été mis à jour par Modgen.

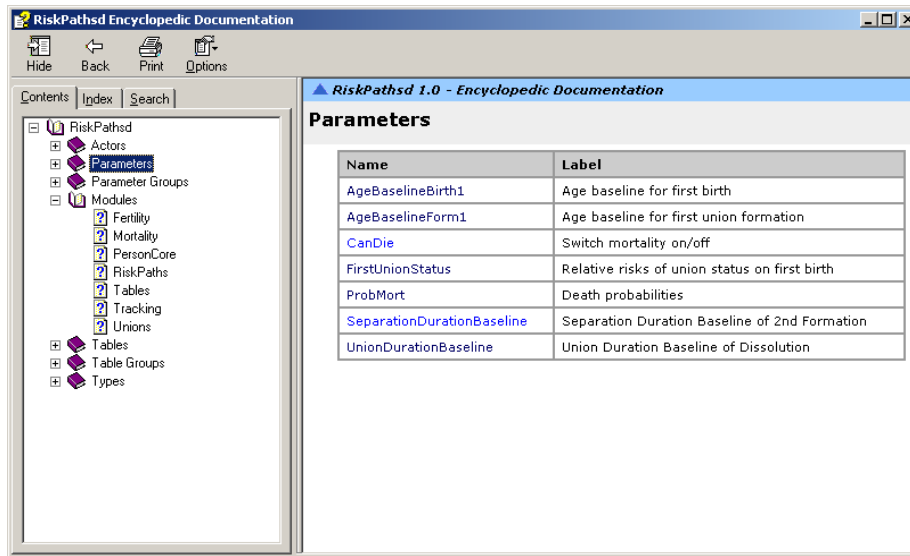
#### **4.4 Sortie de tableau : agrégats et distributions**

Les résultats des simulations sont inscrits dans des tableaux de sortie prédéfinis. Il convient de souligner que les valeurs affichées dans les tableaux de sortie ne représentent qu'une parmi plusieurs vues possibles des résultats. En cliquant sur un tableau avec le bouton de droite de la souris, il est possible d'avoir accès à une fiche de propriétés du tableau. Entre autres, elle permet l'affichage de l'information distributionnelle (les erreurs-types et le coefficient de variation) de toutes les valeurs simulées. Le contenu des tableaux peut également être copié et collé. Vous pouvez choisir de copier le tableau tel qu'il est affiché ou toutes les dimensions du tableau en une seule fois (s'il existe plus de deux dimensions).

#### **4.5 Fonction d'aide et documentation du modèle**

Comme toutes les applications Modgen, RiskPaths fournit des fichiers d'aide de divers types. Deux d'entre eux concernent Modgen proprement dit, à savoir un guide général de l'utilisateur de l'interface visuelle ainsi que des notes sur les versions de Modgen. Les autres fichiers d'aide sont particuliers au modèle. Toutes les applications Modgen contiennent un fichier de documentation encyclopédique détaillé sur le modèle. Cette documentation est créée automatiquement à partir d'un code commenté de manière appropriée.

Figure 3 : Documentation sur le modèle



#### 4.6 Sortie graphique des biographies individuelles

L'application de consultation des biographies (BioBrowser) de Modgen est un outil qui permet l'affichage graphique des trajectoires de vie individuelles. Cette visualisation des résultats des simulations est particulièrement utile pour le débogage du modèle. Afin d'utiliser l'outil, la fonction de dépistage (tracking) doit être activée dans les paramètres du scénario. La liste des variables à dépister doit aussi être déclarée par le développeur du modèle dans le code du modèle au moyen d'une instruction de dépistage. Modgen dépiste alors tous les changements concernant les variables incluses dans l'instruction de dépistage pour un échantillon d'acteurs simulés (dont la taille est spécifiée comme l'un des paramètres du scénario).

Affichez les biographies créées par RiskPaths, lancez simplement l'application BioBrowser et chargez le fichier de dépistage de votre scénario de simulation, p. ex. Base(trk).mdb.



Figure 4 : BioBrowser

